6X-03

動画からのヴィネットイラスト半自動生成

生井 麻結　　藤代 一成
慶應義塾大学 理工学部情報工学科

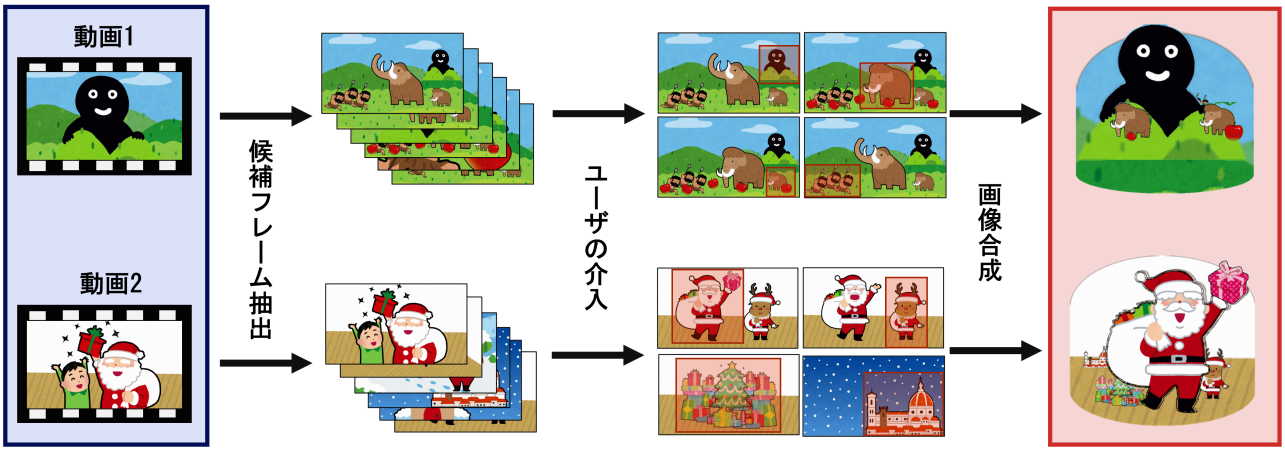


図 1: 提案フレームワークと出力結果．青枠で囲まれた動画 1，2 は入力を表し，赤枠で囲まれた画像は出力結果を表す．

1　背景と目的

現在，肥大化するメディアデータに対処するため，様々な要約技術が発表されている．しかし，その多くは同種メディア変換であり圧縮率に限界があるため，異種メディア変換の可能性が示唆されてきた．そこで本研究では，ヴィネットイラスト (vignette illustrations) に注目する．ヴィネットイラストは，ビデオ映像作品やゲームの物語コンテンツを凝縮して表現し，一瞥でその世界観を理解させられる．動画からヴィネットイラストへ変換するとき，それは異種メディア変換となる．同種メディア変換である既存の要約技術の本質は，要素の選択にある．しかし，動画からヴィネットイラストへの変換は，選択に加えて要素の位置決めや色調の微調整等が必要である．その点から，より挑戦的な要約技術であるといえる．

異種メディア変換の既存の研究として Toyoura らの方法 [1] が存在する．これは動画からフィルムコミックへの変換であるが，本研究は 1 枚のイラストに変換するため，情報の圧縮率がさらに高い．一方，ヴィネットイラストにおけるエフェクト生成の研究として Ikeda らの研究 [2] があるが，これは効果的なエフェクト生成だけに注力しており，ヴィネットイラスト自体を生成するものではない．

本発表では，機械学習を用いて動画からヴィネットイラストへ自動変換する手法の一案を示す．**図** 1 に提案フレームワークとその出力結果の例を示す．

2　ヴィネットイラストとは

Pinterest 等の関連ウェブサイトに掲載されている事例 2 万枚余を事前解析した結果，ヴィネットイラストは以下の 5 種類の要素から構成されていることが分かった．

- キャラクタ：物語の主人公や登場人物
- ステージ：物語が展開される主要舞台
- バックグラウンド：物語を象徴する背景
- サポータ：脇役を務める調度品や小物
- エフェクト：キャラクタを引き立てる視覚効果

図 2 に，ヴィネットイラストとその構成要素への分解例を示す．キャラクタとステージは必須要素である．残り三種は必要に応じて組合せ可能であるが，上に列挙した順に世界観の表現度は高く，逆に演出介入度は低い．本研究では，この事前解析に基づいてヴィネットイラストを生成する．

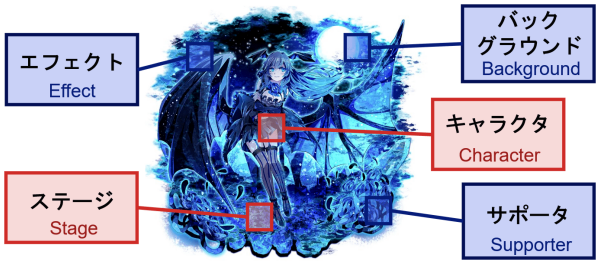


図 2: ヴィネットイラストの例とその構成要素への分解例．

Semiautomatic generation of vignette illustrations from video
Mayu Namai, Issei Fujishiro
Department of Information and Computer Science, Keio University

3　手法

本節では，提案フレームワーク（**図** 1）における各処理について詳述する．基本的なアイディアとしては，まず動画をいくつかのフレーム画像に変換したあとで，段階的に情報の圧縮率を上げ，最終的に一枚のヴィネットイラストへ変換する．

3.1　候補フレーム抽出

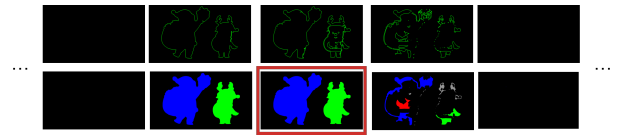
本手法における候補フレームとは，動画のカットのなかで最も代表的なフレーム画像と定義する．今回のフレームワークでは，動体の動きが最も大きいフレームを候補フレームとして抽出した．カット検出は，前後のフレーム間の平均絶対誤差を計算し，閾値を設定することで判定する．動体検知の手法としては，まず前後のフレームを比較して移動平均を計算し，注目しているフレームにおいて変化した部分だけを抜き出す．その後変化した部分を動体とし，動体の境界線を抽出して塗りつぶすことで動体のマスク画像を得る．この処理を動画内のすべてのフレーム画像に施すことで，カット内で動体の動きが最も大きいフレームを抽出できる．**図** 3(a) に，動画 2 を入力とした際の動体検知によるマスク生成の様子を示す．

3.2　ユーザの介入

前項において抽出した候補フレームのなかから，ユーザに気に入ったキャラクタ，サポータが映っているフレームを選択させる．**図** 3(b) に，動画 2 を入力とした際のユーザによる領域選択の様子を示す．マウスを用いてユーザが注目している領域を四角く囲ませ，その物体の透過画像を生成する．その手法として，その四角内のキャラクタ，サポータを動体検知時に得られたマスクによる切り抜き，または Grabcut を用いる．さらに，キャラクタが映っているフレーム画像からユーザにステージの領域を指定させる．この手続きにより，バックグラウンドとステージの透過画像を生成する．実際のヴィネットイラストの特徴にしたがって，バックグラウンドの画像全体にはぼかしをかけている．また，その形状は円形または四角形の二通りの形状が出力され，ユーザは好みの形を選ぶことができる．

3.3　画像合成

取得したキャラクタ，サポータ，ステージ，バックグラウンドの透過画像を重畳し，最終的なヴィネットイラストを生成する．サポータの位置座標とサイズはランダムフォレストを用いて予測する．ランダムフォレストの学習データには，実際のヴィネットイラストから取得した，ヴィネットイラスト全体に占めるステージの位置，キャラクタとサポータの座標，サイズ，縦横比を用いた．ステージの位置，キャラクタの位置，サイズ，縦横比は既に前項の処理時に分かっているので，それらの値をもとに回帰分析することで，サポータに関する情報を予測する．



(a) **動体検知によるマスク生成**

(b) **ユーザによる領域選択**

図 3: **動体検知によるマスク生成とユーザによる領域選択．**(a) **動体を検知しアニメ動画から線画を抽出し，そこからマスク画像を得る．赤枠の画像は動体の動きが最も大きいフレームのマスク画像である．**(b) **その後ユーザの介入により，注目している物体のくり抜き画像を得る．青枠の画像は領域選択時の実際のインタフェースである．**

4　結果

今回は二種類の動画を用いて実験した．**図** 1 に入力動画から抽出されたフレーム画像の一部と出力結果のヴィネットイラストを示す．動画 1 は主に屋外のシーン，動画 2 は屋内のシーンが中心である．また動画の長さは動画 1 が 1 分 2 秒，動画 2 が 1 分 47 秒である．どちらの動画においてもユーザが選択したキャラクタとサポータ，それらに合わせたステージとバックグラウンドによってヴィネットイラストが構成されていることが分かる．

5　結論と今後の課題

本稿では機械学習を用いて，動画からのヴィネットイラストを半自動生成する手法を提案した．内容や長さの異なる二種類の動画を入力し，その内容を含むヴィネットイラストを生成できることを確認した．

今後の課題として，ヴィネットイラストの一要素であるエフェクトの生成が挙げられる．また，動画の内容から要素のサイズ感を把握する仕組みを追加することや，より一般的な動画へ適用可能にしていくことが挙げられる．さらに今後の展望として，生成系拡散モデルの導入や，ユーザの興味をより反映したイラストを生成するために，視線追跡による完全自動生成を検討している．

謝辞

本研究の一部は，令和 4 年度科研費挑戦的研究 (開拓)20K20481 の支援により実施された．

参考文献

[1] Masahiro Toyoura, Tomoya Sawada, Mamoru Kunihiro, and Xiaoyang Mao: “Using eye-tracking data for automatic film comic creation,” in *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12). ACM*, New York, NY, USA, pp. 373–376, 2012.

[2] Riwano Ikeda and Issei Fujishiro: “SpiCa: Stereoscopic effect design with 3D pottery wheel-type transparent canvas,” in *SIGGRAPH Asia 2021 Technical Communications. ACM*, New York, NY, USA, Article 14, pp. 1–4, 2021.